International Journal of of Grid and Distributed Computing
Vol.2, No.2, June 2009

Mobile Queries in GIS

Debnath Bhattacharyya[1], Debasri Chakraborty[1], Poulami Das[1],
Samir Kumar Bandyopadhyay[2], and Tai-hoon Kim[3]

[1]*Computer Science and Engineering Department, Heritage Institute of Technology, Kolkata-700107, India*

[2]*Department of Computer Science and Engineering, University of Calcutta, Kolkata-700009, India*

[3]*University of Tasmania, School of Computing and Information Systems, Hobart, Australia*
{debnathb,chakraborty.debasri,dasp88}@gmail.com[1], skb1@vsnl.com[2]
taihoonn@empal.com[3]

Abstract

In this paper we propose an algorithm for exhaustive search of GIS Mobile query (point-in-area function). This algorithm search the status of a given co-ordinate point with respect to an area (described as a set of co-ordinates without any geometric relationship). In the exhaustive search algorithm proposed previously, there were a number of undetermined points (which does not support the range of scale), which are close to boundary co-ordinates. In this paper a new approach is introduced by which much accurate result can be obtained.

Keywords: *GIS, image, geometric and digital.*

1. Introduction

Geographic Information Systems (GISs) deal with the collection, management, and analysis of large volume of spatially referenced data [1] together with the associated logical and/or numerical attributes (semantics and thematic) [2]. The ability to manipulate these spatial data into different forms with reference to the semantics and/or thematic involved and to extract the additional information from them is the root of GIS technology.

The source of GIS data can be classified into two categories: (1) the coordinates of different features (point, line, area) along with the titles from hand drawn maps, and (2) the logical and numerical information regarding the different objects of the maps from the tables associated with the maps. GIS queries can be done in the same way either it is some logical query or query may be related with the coordinate's features. A land can be defined as an area, road, rivers, rail-lines can be described by line features, a particular object in an area can be defined as a point feature (same as, an area can be defined as a point feature with respect to a vast area.). Queries can be done to find the position of a coordinate feature with respect to another coordinate feature. The coordinate objects may be still or moving.

2. Previous works

The features (point, line, area) are stored in the cartographic database with the primary objective of extracting the desired semantic and thematic information out of these large volumes of geometrically unrelated data. Since, this database, as developed by the digitization method [1], consists of very large number of co-ordinates of points without any established geometrical relationship among them. The results of Euclidean geometry are not applicable to this case. Further the necessity of very high degree of accuracy makes the conventional approaches [1] inadequate for design of a cartographic database.

The design of any GIS is largely dependent on two major components, such as, (1) the data organization for storage and retrieval of spatial and aspetial data related to geographic objects (features), and (2) the query processor working as an interface between the data organization and user defined queries.

The performance of any query processor is very much dependent on the respective data organization. The GEOQUEL, a special purpose language, proposed by Berman and Stonebaker [5], was one of the earlier attempts to make use of an existing DBMS for GIS data processing. This scheme decomposes every line/area feature into a sequence of line segments. It does not support geographic data types and hence, retrievals based on spatial relationships are not possible.

An attempt to extend SQL for geographical application was also made by A. Frank in the proposed MAPQUERY [6] system. This is a land information system with extra constructs to support geographical input and output facilities. It also does not support spatial operators. Pictorial SQL (PSQL) [7, 8] has been proposed as an interface language for retrieving data from pictorial database with two additional options ON and AT for selecting an area of a given picture.

B. C. Ooi in his activities [4] has adopted this approach to form a GEOgraphic Query Language (GEOQL), which is an extension of SQL. This research work is based on a special data model, named skd-tree with the additional GEOQL operators intersection, adjacent, ends at, etc.

The speed of processing of GIS queries can be enhanced by designing the algorithms to suit a parallel platform [9-11]. In fact parallelism is inherent in GIS processing though there exists no such significant effort for exploring the aspect. As the stored data does not possess any intrinsic geometrical properties, the processing of GIS queries are not at all data dependant, and, due to the large volume of data to be handled for any GIS query, application of parallel algorithms has a very significant effect [12, 13] on the processing of GIS queries.

R. Dasgupta in his research activities [16] proposed a general purpose GIS system for answering a wide spectrum of queries in a suitable parallel platform for higher efficiency. The concept of an address square has been utilized for improving the efficiency of the system. Some terminologies such as Geographic Information Unit (GIU), Address Square, Boundary Address Square (BAS), Internal Address Square (IAS) is used, where GIU is a cartographic information which is divided into some n n grid structure and each grid was defined as address square. An area is defined as a set of co-ordinate points and in which square the boundary point lies is said BAS, which address square is entirely inside the area is said IAS.

R. Dasgupta had proposed some algorithms to query of some basic GIS functions (such as, Point-In-Area, Line-In-Area, and Area-In-Area) where significantly less amount of exhaustive computation is done.

3. Our work

We propose our work by considering the following points:

3.1. Data capture

How can a GIS use the information in a map? If the data to be used are not already in digital form, that is, in a form the computer can recognize, various techniques can capture the information. Maps can be digitized by hand-tracing with a computer mouse on the screen or on a digitizing tablet to collect the coordinates of features. Electronic scanners can also convert maps to digits. Coordinates from Global Positioning System (GPS) receivers can also be uploaded into a GIS.

Data capture — putting the information into the system involves identifying the objects on the map, their absolute location on the Earth's surface, and their spatial relationships. Software tools that automatically extract features from satellite images or aerial photographs are gradually replacing what has traditionally been a time-consuming capture process. Objects are identified in a series of attribute tables the "information" part of a GIS. Spatial relationships, such as whether features intersect or whether they are adjacent, are the key to all GIS-based analysis.

3.2. Data integration

A GIS makes it possible to link, or integrate, information that is difficult to associate through any other means. Thus, a GIS can use combinations of mapped variables to build and analyze new variables (Figure-1).

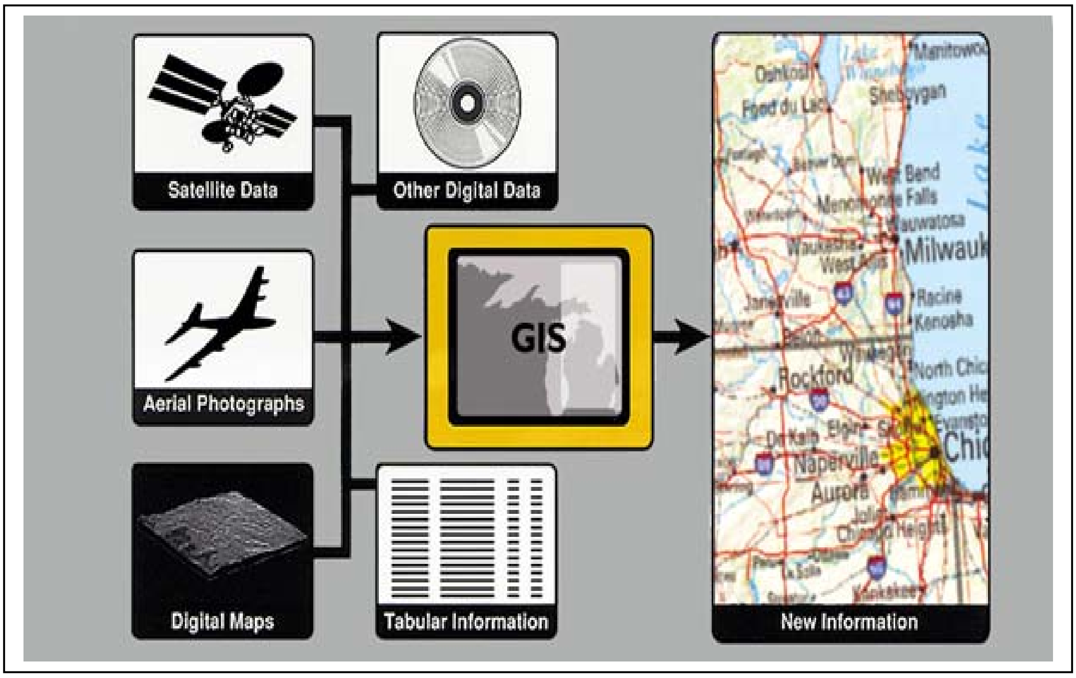


Figure 1. Data integration is the linking of information in different forms through a GIS

For example, using GIS technology, it is possible to combine agricultural records with hydrograph data to determine which streams will carry certain levels of fertilizer runoff. Agricultural records can indicate how much pesticide has been applied to a parcel of land.

Locating these parcels and intersecting them with streams can use the GIS used to predict the amount of nutrient runoff in each stream. Then as streams converge, the total loads can be calculated downstream where the stream enters a lake.

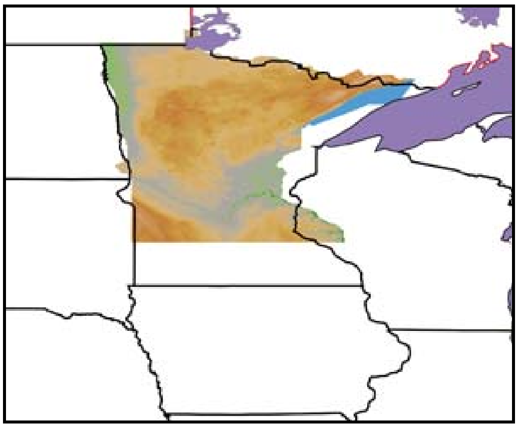


Figure 2. An elevation image classified from a satellite image of Minnesota exists in a different scale and projection than the lines on the digital file of the State and province boundaries

3.3. Projection and registration

A property ownership map might be at a different scale than a soils map. Map information in a GIS must be manipulated so that it registers, or fits, with information gathered from other maps. Before the digital data can be analyzed, they may have to undergo other manipulations projection conversions, for example that integrate them into a GIS.

Projection is a fundamental component of mapmaking. A projection is a mathematical means of transferring information from the Earth's three-dimensional, curved surface to a two-dimensional medium paper or a computer screen. Different projections are used for different types of maps because each projection is particularly appropriate for certain uses. For example, a projection that accurately represents the shapes of the continents will distort their relative sizes.

Since much of the information in a GIS comes from existing maps, a GIS uses the processing power of the computer to transform digital information, gathered from sources with different projections, to a common projection (Figure 2 and Figure 3).

3.4. Data structure

Can a land use map be related to a satellite image, a timely indicator of land use? Yes, but because digital data are collected and stored in different ways, the two data sources may not be entirely compatible. Therefore, a GIS must be able to convert data from one structure to another.

Satellite image data that have been interpreted by a computer to produce a land use map can be "read into" the GIS in raster format. Raster data files consist of rows of uniform cells coded according to data values. An example is land cover classification (Figure 4). The computer can manipulate raster files quickly, but they are often less detailed and may be less visually appealing than vector data files, which can approximate the appearance of more traditional hand-drafted maps. Vector digital data have been captured as points, lines (a series of point coordinates), or areas (shapes bounded by lines) (Figure 5). An example of data typically held in a vector file would be the property boundaries for a particular housing subdivision.

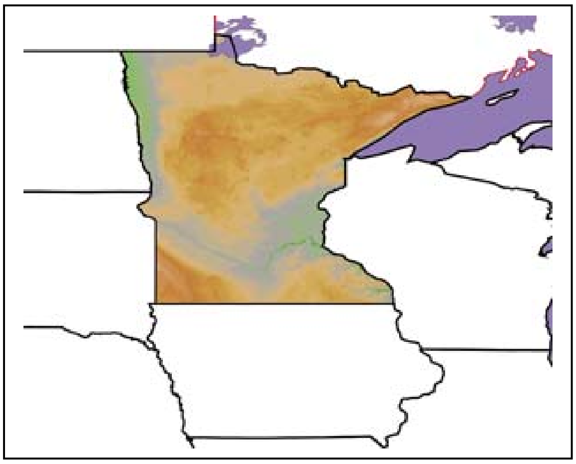


Figure 3. The elevation image re-projected to match the projection and scale of the State and province boundaries

1	1	1	1	1	1	1	3	3	3
1	1	1	1	1	1	1	3	3	3
1	1	1	1	1	1	3	3	3	3
1	1	1	2	2	2	2	3	3	3
1	1	1	2	2	2	2	3	3	3
1	1	1	2	2	2	2	3	3	3
1	1	1	1	2	2	2	3	3	3
1	1	1	1	1	1	3	3	3	3
1	1	1	1	1	1	1	3	3	3
1	1	1	1	1	1	1	1	3	3

Figure 4. Example of the structure of a raster file

3.5. Data modeling

It is impossible to collect data over every square meter of the Earth's surface. Therefore, samples must be taken at discrete locations. A GIS can be used to depict two- and three-dimensional characteristics of the Earth's surface, subsurface, and atmosphere from points where samples have been collected.

3.6. Algorithm for exhaustive search (Sum angle approach)

INPUT: All boundary co-ordinate points of the area feature and the coordinate of point feature to be searched.

OUTPUT: The status of the co-ordinate with respect to the area feature.

PROCEDURE: Suppose there are ‘n’ numbers of boundary point co-ordinates present for an area.

Step 1: Pass the co-ordinate of the test point feature.
Let it be (x_p, y_p).
Step 2: Pass all the co-ordinates of the area feature.
Step 3: For all the area points calculate

$$\sum_{i=1}^{n} \angle (x_i, y_i)\,(x_p, y_p)\,(x_{i+1}, y_{i+1})$$

Where, $(x_i, y_i)\,(x_p, y_p)\,(x_i+1, y_i+1)$ represents the angle formed by (x_i, y_i) and (x_i+1, y_i+1) with (x_p, y_p).

Step 4: calculate the direction of the angle either it is clockwise or anticlockwise.
Step 5: subtract the angle if it is clockwise and add the angle if it is anticlockwise. Finally find the angle made by the point with the area.
Step 6: If

$$\sum_{i=1}^{n} \angle (x_i, y_i)\,(x_p, y_p)\,(x_{i+1}, y_{i+1}) = 2\pi$$

then, the point is inside.

If

$$\sum_{i=1}^{n} \angle (x_i, y_i)\,(x_p, y_p)\,(x_{i+1}, y_{i+1}) = 0$$

then, the point is outside.

The algorithm described above for exhaustive search is the sum angle approach which says that we have to calculate the angle between each two consecutive points and the given searched point and sum up those small angles to find the angle formed by the searched point with area. The coordinate points of the area need to be stored anticlockwise fashion as per the algorithm. If the point is inside then the ultimate angle would be 360^0 as there is no direction change but if the point is outside the area then the ultimate sum of the angles would be 0^0 as there is a direction change in time of calculating the angle.

4. Result

Previously proposed algorithm has been implemented by us, and the result is shown in Table 1. Where, undetermined points are determined as those points those are not supporting the range. The ranges are determined with tolerance value of 355^0 to 365^0 (-5 to +5 of 360^0 for inside) and -5^0 to $+5^0$ (-5 to +5 of 0^0 for outside). It is noticed that the coordinate points those are giving abnormal results are very close to boundary points.

If the coordinate points are on the boundary then it will also give out of range exception and that is also undetermined.

Result of our algorithm is shown in Table 2. Here, undetermined points are determined as those points those are not supporting the range. The ranges are determined without tolerance value that of the previous (360^0 for inside and 0^0 for outside). It is noticed that the coordinate points those are giving abnormal result are only on boundary.

Table 1. Results of previously proposed exhaustive search algorithm

No of Boundary points of the area	No of randomly generated points for test	No of points that are inside	No of points that are outside	No of points that remain undetermined
312	1000	393	498	109
503	1000	762	216	22
386	1000	423	501	86

Table 2. Result of our exhaustive search algorithm.

No of Boundary points of the area	No of randomly generated points for test	No of points that are inside	No of points that are outside	No of points that remain undetermined
312	1000	452	548	0
503	1000	762	238	0
386	1000	449	551	0

5. Conclusion

Both the algorithms for exhaustive searching are implemented in VC++ (greater size, minimum 512MB, of the RAM will require). The three areas are taken with the coordinate points of number 312, 503, 386 for boundary coordinate values. Test points are selected randomly with some user given range. We have taken 1000 random points for testing and results are analyzed above. It can be noticed that if we calculate direction of the angle manually (which was not proposed) then it will give a better result as the same domain values are there and the result is more accurate as there is no tolerance value as before (angle is exact 3600 and 00). The points on the boundary are still remains undetermined. Thus we can say that the proposed algorithm is giving better result than previous. Experiments with greater domain remain as future work.

References

[1] R. Kasturi and J. Alemany, “Information extraction from images of paper-based maps”, IEEE Transactions on Software Engineering, Vol. 14, Issue 5, May 1988, pp. 671-675.

[2] B. B. Pal and S. Bhattacharya, “Development of an efficient data structure for solving semantic queries related to cartographic database”, International Conference On Database, Parallel Architectures and Their Applications, 7-9 March, 1990, Miami Beach, FL, USA, pp. 100-102.

[3] T. Matsuyama, L. V. Hao and M. Nagao, “A file organization for geographic information systems based on spatial proximity”, International Journal Computer Vision, Graphics and Image Processing, June 1984, Vol. 26, No. 3, pp. 303-318.

[4] Beng Chin Ooi, “Efficient query processing in geographic information systems”, Lecture notes in Computer Science, Springer-Verlag New York, USA, Vol. 471, pages: 208.

[5] R. Berman and M. Stonebraker, “GEO-QUEL: A system for manipulation and display of geographic data”, ACM Computer Graphics, Vol. 11, No. 2, 1977, pp. 186-191.

[6] A. Frank, “Mapquery: database query language for retrival of geometric data and their graphical representation”, ACM SIGGRAPH Computer Graphics, July 1982, Vol. 16, Issue 3, pp. 199-207.

[7] N. Roussopoulos and D. Leifker, “An introduction to PSQL: a pictorial structured query language”, IEEE Workshop on Visual Languages, December 1982, Hirosima, Japan, pp. 77-87.

[8] N. Roussopoulos and D. Leifker, “Direct spatial search on pictorial databases using packed R-trees”, ACM SIGMOD, International Conference on management of data, May 1985, Austin, Texas, USA, Volume 14 , Issue 4, pp. 17-31.

[9] D. L. Eager, J. Zahorjan and E. D. Lazowska, “Speedup versus efficiency in parallel systems”, IEEE Transactions on Computers, March 1989, Vol. 38, No. 3, pp. 408-423.

[10] David C. Fisher, “Your favorite parallel algorithm might not be as fast as you think”, IEEE Transactions on Computers, February 1988, Vol. 37, Issue 2, pp. 211-213.

[11] M. J. Quinn, “Designing efficient algorithm for parallel computers”, January 1987, McGraw-Hill College, New York, USA, pages: 286.

[12] R. Dasgupta, B. B. Pal and S. Bhattacharya, “Design of an efficient parallel algorithm for solving semantic queries related to cryptographic database”, SEARCC-90, December 1990, Manila, Philippines.

[13] R. Dasgupta, S. Bhattacharya and S. Bandyopadhyay, “Maching of large files: a parallel algorithmic approach”, National Conference on Information Technology, CSI90, October-November 1990, Calcutta, India.

[14] R. Dasgupta, S. Bhattacharya, “A parallel for integration of cartographic data related to integrated geographic information system”, National Seminar on parallel computer systems and their applications, October 1990, Calcutta, India.

[15] R. Dasgupta and S. Bhattacharya, “Design of an efficient query processor for Geographic Information”, International Conference on Applied Informatics, May 1993, Annecy, France.

[16] R. Dasgupta, N. Chaki and S. Bhattacharya, “Design of an Integrated Geographic Information System: An Algorithmic Approach”, International Conference on Intelligent information System, June 1996, Washington D.C., USA.

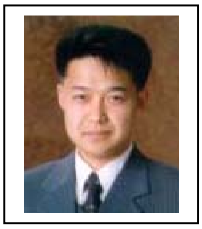
Authors

Debnath Bhattacharyya, M.Tech in Computer Science and Engineering from West Bengal University of Technology, Kolkata. He is working as a Faculty Member with the Computer Science and Engineering Department at Heritage Institute of Technology, Kolkata. He has 14 years of experience in Teaching and Administration. His research interests include Bio-Informatics, Image Processing and Pattern Recognition. He has 14 Years of experience in the line of Teaching and Projects. He is working towards his research, since, middle of 2006 under the supervisions of Prof. Samir Kumar Bandyopadhyay, University of Calcutta, Kolkata. He has published 45 Research Papers in International Journals and Conferences and 3 Text Books for Computer Science.

Debasri Chakraborty, M.Tech in Computer Science and Engineering with specialization in Multimedia and Software Systems from University of Technology. She is currently a Lecturer with the Computer Science and Engineering Department at Heritage Institute of Technology. Her research interests includeGIS, Image Processing and Pattern Recognition. She has 3 years of experience in teaching and research. She has published one research paper in a National Conference.

Poulami Das, M.Tech in Computer Science and Engineering from the University of Calcutta. She is currently a Lecturer with the Computer Science and Engineering Department at Heritage Institute of Technology, Kolkata. Her research interest include Bio-Informatics and Image Processing. She has 5 Years of experience in the line of Teaching. She is working towards his research, since, middle of 2006 under the guidance of Prof. Samir Kumar Bandyopadhyay. She has published 38 Research Papers in International Journals and Conferences and 1 Text Book for Computer Science.

Dr. Samir Kumar Bandyopadhyay, B.E., M.Tech., Ph. D (Computer Science & Engineering), C.Engg., D.Engg., FIE, FIETE, currently, Professor of Computer Science & Engineering and Registrar, University of Calcutta, visiting Faculty Dept. of Comp. Sc., Southern Illinois University, USA, MIT, California Institute of Technology, etc. His research interests include Bio-medical Engg, Mobile Computing, Pattern Recognition, Graph Theory, Software Engg.,etc. He has 25 Years of experience at the Post-graduate and under-graduate Teaching & Research experience in the University of Calcutta. He has already got several Academic Distinctions in Degree level/Recognition/Awards from various prestigious Institutes and Organizations. He has published 300 Research papers in International & Indian Journals and 5 leading text books for Computer Science and Engineering. He has visited USA, Finland, Sri Lanka.



Prof. Tai-hoon Kim, M.S., Ph. D (Electricity, Electronics and Computer Engineering), currently, Professor of Hannam University, Korea. His research interests include Multimedia security, security for IT Products, systems, development processes, operational environments, etc. He has 14 Years of experience in Teaching & Research. He has already got distinctive Academic Records in international levels. He has published more than 100 Research papers in International & National Journals and Conferences.